Information Engineering (IE) Volume 5, 2016 www.seipub.org/ie
doi: 10.14355/ie.2016.05.001

Video Denoising Based on Riemannian Manifold Similarity and Rank-One Projection

Yongchao Wang*1, Xiaohua Li2,Lei Zhao3

*1,2Information Center of Zhejiang University

3 The Department of Computer and Science, Zhejiang University, 310027, China

*1Ychwang@zju.edu.cn; 2xiaohua@zju.edu.cn,3cszhl@zju.edu.cn

Abstract

In this paper, we combine two powerful tools to handle the video denoising problem: one is an effective video denoising method based on Riemannian Manifold Similarity, and the other is a Rank-One Projection matrix completion based on video denoising method. Similarly, in our algorithm, a noisy video is processed in block-wise manner for each processed block. we form a 3D data array that we call "group" by stacking together blocks which is found similar to the currently processed one. "Collaborative filtering" exploits the correlation between grouped blocks and the corresponding highly sparse representation of the true signal in the transform domain. By employ Rank-One Projection matrix completion method in our framework, our technique is also robust to different types of noise. Experiments demonstrate that our techniques produce state-of-the-art results for video denoising applications.

Keywords

Video Denoising, Rank-One Projection Matrix Completion, Riemannian Manifold Similarity, Block Matching

Introduction

With today's advances in sensor design, the image/video is relatively clean for high-end digital cameras at low sensitivities, but it remains noisy for low cost cameras at high sensitivities, e.g., low light condition, high ISO setting and high speed rate. The problem of removing image noise is still of acute and in fact growing importance with the prevalence of webcams and mobile phone cameras. A recent denoising strategy, the non-local spatial estimation [2], has also been adapted to video denoising [3]. In this approach, similarity between 2D patches is used to determine the weights in a weighted averaging between the central pixels of these patches. For image denoising, the similarity is measured for all patches in a 2D local neighborhood centered at the currently processed coordinate. For video denoising, a 3D such neighborhood is used. The effectiveness of this method depends on the presence of many similar true-signal blocks[4].Based on the same assumption as the one used in the non-local estimation, i.e. that there exist mutually similar blocks in natural images, in [5] the authors proposed an image denoising method. There, for each processed block, we perform two special procedures: grouping and collaborative filtering. Grouping finds mutually similar 2D blocks and then stacks them together in a 3D array that we call group. The benefit of grouping highly similar signal fragments together is the increased correlation of the true signal in the formed 3D array. Collaborative filtering takes advantage of this increased correlation to effectively suppress the noise and produces estimates of each of the grouped blocks. They showed [5] that this approach is very effective for image denoising.In this paper, we apply the concepts of grouping and collaborative filtering to video denoising. Grouping is performed by a specially developed predictive-search block matching technique that significantly reduces the computational cost of the search for similar blocks. We employ a two-step video-denoising algorithm proposed in [4] where the predictive search block-matching is combined with collaborative hard thresholding in the first step and with collaborative Wiener filtering in the second step. In order to enhance the robustness of the algorithm in processing denoising problems with multiple sources of noises, our algorithm is derived with minimal assumptions on the statistical properties of image noise. The basic idea is to convert the problem of removing noise from the stack of matched patches to a low rank matrix completion problem, which can be efficiently solved by minimizing the nuclear norm of the matrix with linear constraints.[1]

Related Work

There has been an abundant research literature on image denoising methods. In this section, we will only discuss the most related denoising techniques. There has been abundant research literature on image denoising methods. In this section, we will only discuss the most related denoising techniques [4,6,7]. Although differing from details, these method are built on the same methodology which essentially groups the similar patches together followed by a collaboratively filtering. Take the well-known BM3D [5] as a sample. In BM3D, similar image blocks are stacked in a 3D array based on the L2 norm distance function between different patches. Then a shrinkage in 3D transform domain such as wavelet shrinkage or Wiener filter is applied on the 3D block stack. The denoised image is then synthesized from denoised patches after inversing 3D transform. The result can be further improved by iteratively doing grouping and collaboratively filtering. Video denoising is different from single image denoising as video sequences usually have very high temporal redundancy which should be effectively used for better performance (e.g., [8-10]). The basic idea of patch-based image denoising can also be applied on the video by matching similar patches both within the image and over multiple images. The concept of BM3D is generalized to video denoising in [4] by using a predictive search block-matching over time and combined with collaborative Wiener filtering on patch stacks. In [11], a more robust patch matching is proposed by using the depth as a constraint in the matching process and the patch stack is denoising by both PCA (principle component analysis) and tensor analysis. The idea of sparse coding in a patch dictionary has also been applied on video denoising, where the denoised image patches are found by seeking for the sparsest solution in a patch dictionary. Among these patch-based video denoising techniques, most assume data noise is only additive i.i.d. Gaussian noise (e.g., [4]). The image noise mixed with both Gaussian noise and Poisson shot noise are considered in [14].

Our Approach

Image comparison is a topic that has received a lot of attention in themage processing and computer vision communities since it is a main ingredient in manyapplications, such as object recognition, stereo vision, image interpolation, image denoising,and exemplar-based image inpainting, among others. A common way to define a nonlocalsimilarity measure between two images is to compare the patches (local neighborhoods) aroundeach pair of points formed by taking one point from each image. We consider a general settingin which images are defined on Riemannian manifolds. Such manifolds arise, for instance, forimages defined on R N , endowed with a suitable metric depending on the image.In [3], it was shown that multiscale analyses of similarities between images on Riemannianmanifolds, satisfying a certain set of axioms, are (viscosity) solutions of a family of degeneratePDEs. Our goal in this paper is to study one particular instance of the set of models derivedin [3], namely a linear model to compare patches defined on two images in R N endowedwith some metric. Except for its genericity, this linear model is selected by its computationalfeasibility since the solution of the PDE can be approximated via the convolution with ashort-time space-varying kernel, leading to an algorithm that has the complexity of the usualEuclidean patch comparison.Let us review the fundamentals of the approach in [3]. Given two images u,v defined intheir respective image domains (assume R 2 for simplicity), we want to compare their neighborhoods at the points x,y ∈ R 2 , respectively. The simplest way to compare them would be to compare the two neighborhoods of x,y using the Euclidean distance. That is, let us define

$$D(t,x,y) = \int_{\mathbb{R}^2} g_t(h)(u(x + h) - v(y + h))^2 \, dh, \tag{1}$$

where g_t is a given windowing function that we assume to be Gaussian of variance t. Thisformula gives an explicit comparison and assumes that the image domain is the Euclideanplane. It generalizes approaches to patch comparison applied, for example, in [8].

In this part, we employ the strategy utilized in [4] to solve our problems. In order to efficiently capture blocks that are part of objects which move across subsequent frames, we propose theuse of linear multiscale analysis of similarities between images on Riemannian manifold which finds similar (matching) blocks by searching in a dataadaptive spatio-temporal subdomain of the video sequence.For a given reference block located at $x = (x_1, x_2, t_0)$,

whenusing a temporal window of $2N_{FR}+1$ frames.

For each group formed in the former stage, a patch matrix is constructed by concatenating every patch in the group into a long vector and stacking all the vectors. With the purpose of restoring the patch matrix, we adopt a similar method discussed in [1]. For each patch, similar patches are found in both spatial andtemporal domain by using the patch matching algorithm described in the previous section to form the matrix $P_{j,k}$. Theset of missing elements of $P_{j,k}$ have two subsets: the firstpart covers those pixels corrupted by impulsive noise usingthe adaptive median filter based impulsive noise detector([9]). The second partincludes the pixels whose valuediffers from the mean of the corresponding row vector bythe amount larger than a pre-defined threshold. Then Ω isformed by including the index of all remained pixels. As described in [1], $Q_{j,k}$ is recovered fromits incomplete observation $P_{j,k}|_\Omega$ by solving the followingminimization problem:

$$\begin{aligned}&\min_Q \|Q\|_* \\ &s.t.\ \|Q|_\Omega - P|_\Omega\|_F^2 \le \#(\Omega)\hat{\sigma}^2,\end{aligned} \tag{2}$$

where $\hat{\sigma}$ is the estimate of standard deviation of noise, which is obtained by calculating the average of the variances of all elements $\in\Omega$ on each row.

we introduce a "Rank-One Projection" (ROP) model for low-rankmatrix recovery and a constrained nuclear norm minimization method forthis model. Under the ROP model, we observe

$$y_i = (\omega^i)^T A \partial^i + z_i, \quad i=1,.......,n, \tag{3}$$

where ω^i and ∂^i are random vectors with entries independently drawn fromsome distribution P , and z_i are random errors. In terms of the linear map X :R $^{p1\times p2}\rightarrow$ R ^{n} in (1.1), it can be defined as

$$[X(A)]_i = (\omega^i)^T A \partial^i, \ i=1,.......,n, \tag{4}$$

Since the measurement matrices X i = ω^i A $(\partial^i)^T$ are of rank-one, we call themodel(3)a"Rank-OneProjection"(ROP)model.Itiseasytoseethatthestoragefor the measurement vectors in the ROP model (3) is O(n(p 1 +p 2)) bytes whichis significantly smaller than O(np 1 p 2) bytes required for the Gaussian ensemble.We first establish a sufficient identifiability condition in Section 2 by considering the problem of exact recovery of low-rank matrices in the noiseless case. It isshown that, with high probability, ROP with n 〉 r(p 1 +p 2) random projections issufficient to ensure exact recovery of all rank-r matrices through the constrainednuclear norm minimization. The required number of measurements O(r(p 1 +p 2)) is rate optimal for any linear measure ment model since a rank-r matrix A ∈ R p 1 +p 2 has the degree of freedom r(p 1 + p 2 − r). The Gaussian noise case is of particular interest in statistics. We propose a new constrained nuclear norm minimizationestimator and investigate its theoretical and numerical properties in the Gaussiannoise case. Both upper and lower bounds for the estimation accuracy under theFrobenius norm loss are obtained. The estimator is shown to be rate-optimal whenthe number of rank-one projections satisfies either n〉 (p 1 + p 2)log(p 1 + p 2) orn ~ r(p 1 + p 2). The lower bound also shows that if the number of measurementsn < r max(p 1 ,p 2), then no estimator can recover rank-r matrices consistently.The general case where the matrix A is only approximately low-rank is also considered. The results show that the proposed estimator is adaptive to the rank rand robust against small perturbations. Extensions to the sub-Gaussian design andsub-Gaussian noise distribution are also considered.

Experiments and Conculusion

We applied our proposed denoising method on severalvideos with different mixed noise levels. The results arecompared withthat of one existing video denoising method VBM3D [4]. The same algorithm parameters were set as recommended in [4]. It is obvious that theproposed method shows superior preservation of fine imagedetails and at the same time it introduces significantly lessartifacts.In this work, we combine two powerful tools to handle the video denoising problem: one is an effective video denoising method based on Riemannian Manifold Similarity, and the other is a Rank-One Projection matrix completion based on video denoising method. Similarly, in our

algorithm, a noisy video is processed in block-wise manner and for each processed block.This work is supported in part by 973 projects(2012CB725305) and the National Key Technology R&D Program projects (2012BAH70F02, 2013BAH27F03).

REFERENCES

[1] Ji, H., Liu, C., Shen, Z., Xu, Y, “Robust video denoising using low rank matrixcompletion,” CVPR 2010.

[2] A. Buades, B. Coll, and J. M. Morel, “A review of imagedenoising algorithms, with a new one,” Multisc. Model.Simulat., vol. 4, no. 2, pp. 490-530, 2005.

[3] A. Buades, B. Coll, and J. Morel, “Denoising imagesequences does not require motion estimation,”in Proc.IEEE Conf. Adv. Video Signal Based Surveil., AVSS,Palma de Mallorca, Spain, September 2005, pp. 70-74.

[4] Kostadin, D., Alessandro F., and Karen E., “Video denoising by sparse 3D transform-domain collaborative filtering,” EUSIPCO 2007.

[5] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, “Image denoising by sparse 3D transform-domaincollaborative filtering”, IEEE Trans. Image Process.,vol. 16, No. 8, 2007.

[6] M. Elad and M. Aharon, “Image denoising via sparse andredundant representations over learned dictionaries,” IEEETrans. Image Processing, 15(12), 2006.

[7] A. Foi, V. Katkovnik, and K. Egiazarian, “Pointwise shapeadaptive DCT for high-quality denoising and deblocking ofgrayscale and color images,” IEEE Trans. Image Processing,16(5), 2007.

[8] Chen and C. K. Tang, “Spatio-temporal markov randomfield for video denoising,” CVPR 2007.

[9] H. Hwang and R. A. Haddad, “Adaptive median filters: Newalgorithms and results,” IEEE Trans. on Image Processing,4:499–502, 1995.

[10] V. Vaish, M. Levoy, R. Szeliski, C. L. Zitnick, and S. B.Kang, “Reconstructing occluded surfaces using syntheticapertures: Stereo, focus and robust measures,” CVPR 2006.

[11] L. Zhang, S. Vaddadi, H. Jin, and S. K. Nayar, “Multiple viewimage denoising,” ICCV 2009.

[12] D. Rusanovskyy, K. Dabov, and K. Egiazarian,“Moving-window varying size 3D transform-based videodenoising,” in Proc. Int. Workshop on Video Process.and Quality Metrics, Scottsdale, Arizona, USA, January 2006.

[13] C.-C. Wan and C.-H. Hsieh, “An efficient fractal imagecoding method using interblock correlation search,”IEEE Trans. Circuits Syst. Video Tech., vol. 11, no. 2,pp. 257-261, February 2001.

[14] E. Balster, Y. Zheng, and R. Ewing, “Combined spatial and temporal domain wavelet shrinkage algorithmfor video denoising,” IEEE Trans. Circuits Syst. VideoTech., vol. 16, no. 2, pp. 220ñ230, February 2006.

[15] Lin, Z., Chen, M., Ma, Y, “The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices,” UIUC Technical ReportUILUENG-09-2215, 2010.

[16] M. Protter and M. Elad. “Image sequence denoising via sparseand redundant representations,” In IEEE Trans. Image Processing, 2009.